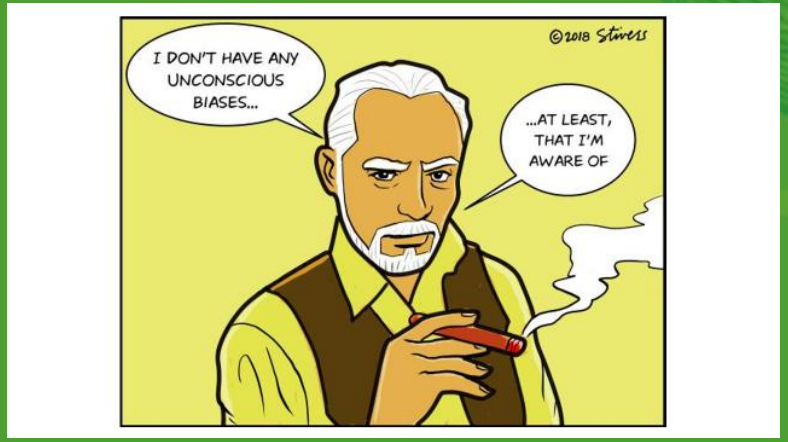
Assessment of biases using distinct metabarcoding technologies for sequencing of fungal ITS amplicons

Carles Castaño[1], Anna Berlin[1], Mikael Brandström[1], Katharina Ihrmark[1], Björn Lindahl[2], Jan Stenlid[1], Karina E. Clemmensen[1]*, Åke Olsson[1]*

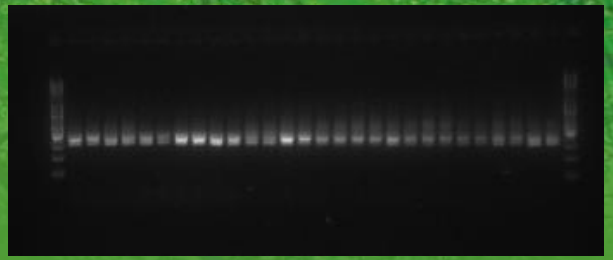
[1] Swedish University of Agricultural Sciences, Department of Forest Mycology and Plant Pathology, SE-75007 Uppsala, Sweden
[2] Department of Soil and Environment, Swedish University of Agricultural Sciences, SE-75007, Uppsala, Sweden.
*shared last authorship

Swedish University of Agricultural Sciences

Liam Klenk



Why optimization of current protocols?



- Most biases occurring during PCR well known. What about sequencing biases?

- Quantify the relative importance of biases (e.g. sequencing, PCR) and correct/account for such biases

- Some studies suggest that HTS data should not be analyzed as relative proportions but as binary data instead…. Also suggested by several reviewers ☹

- Scientific community is sometimes aware of some/most of the biases, but still several protocols are not optimized to minimize the risk of biases

- <u>May affect our results… and ecological interpretations?</u>

Still several protocols are not optimized to minimize the risk of biases

polymerase:DyNAZyme II DNA polymerase, 1:24) and 1 µ
conditions were 94 °C for 5 min; 35 cycles of 94 °C for 1 n
for 30 s, followed by 72 °C for 10 min for primers eub530f
5 min; 35 cycles of 94 °C for 1 min, 55 °C for 1 m

0.2 µl of 0.056 U fast StartExp*Taq* Polymerase (Roche Applied Sciences, Indianapolis, IN, USA), 2.5 µl dNTP (2 mM each), 0.25 µl of each primer, and 1.0 µl of DNA template. Thermocycling conditions were as follows: denaturing at 95 °C for 5 min followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 53 °C for 30 s, extension at 72 °C for 60 s followed by a final extension at 72 °C for 10 min. As negative control, water was used instead of

ecular grade water, 0.5 µl of 10 µm of the 515F, 0.5 µl of 10 µm 806R
The PCR conditions were the following: 94°C for 5 min, followed by 35
llowed by a final elongation of 10 min at 72°C. The composition of the
. The PCR master mix contained 17.1 µl molecular grade water, 1 µl 5 µM

primer, 0.5 µM of the gITS7 primer, and 1U of 1× Phusion High Fidelity DNA Polymerase (New England Biolabs). Thermocycling conditions were as follows: 98 °C for 30 s, 35 cycles of 98 °C for 10 s, 56 °C for 30 s, 72 °C for 15 s and a final extension at 72 °C for 7 min. Negative controls, consisting of sterile water in place of template DNA, did not yield

ately 250 bp) of the nuclear ribosomal rDNA repeat, using the following
one cycle of 95°C for 5 min; then 37 cycles of 95°C for 20 s, 56°C for 30 s and
ending with one cycle of 72°C for 7 min. ITS is the universal DNA
ungi and has been used in a wide variety of taxonomic and ecological

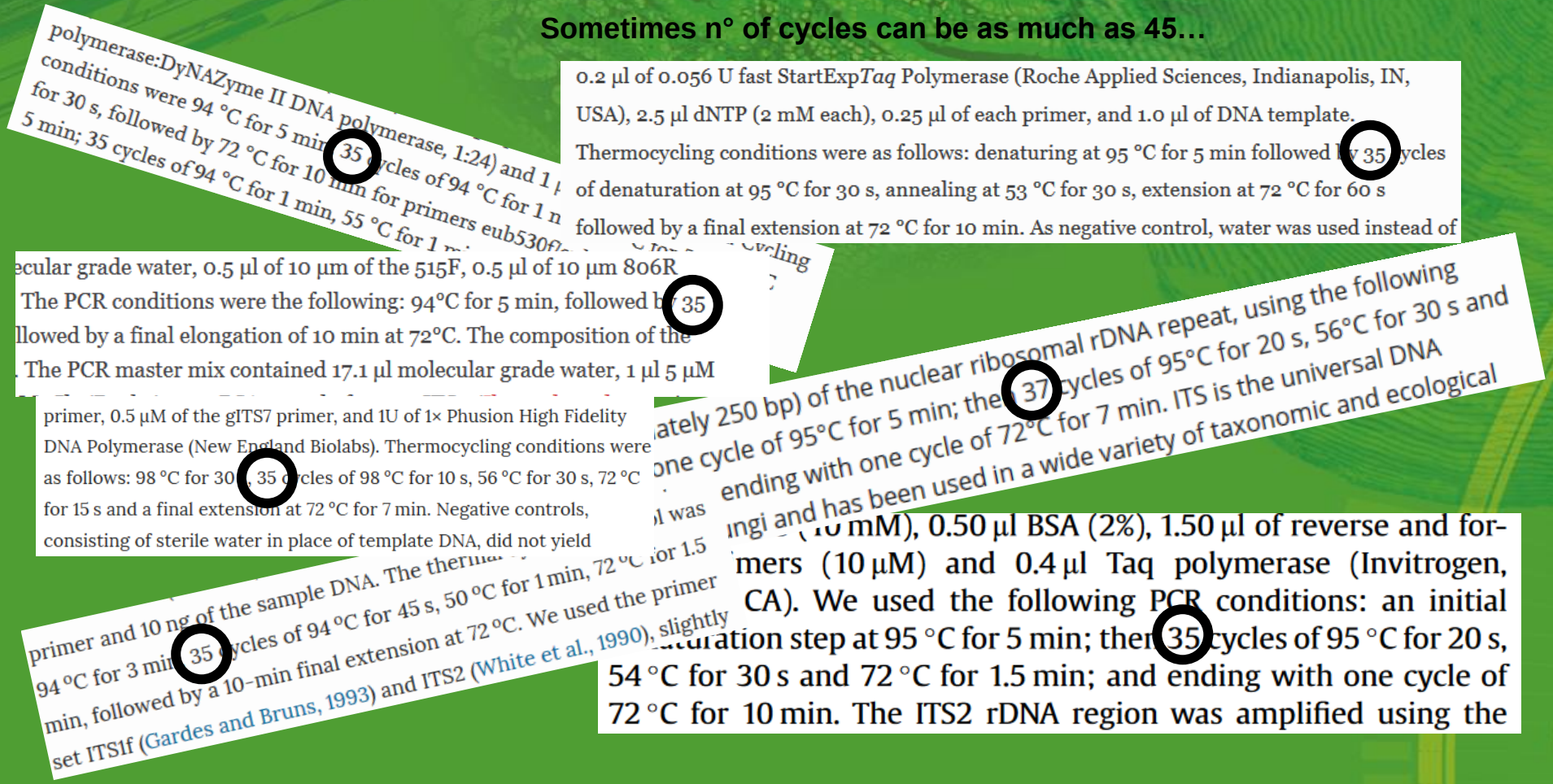
primer and 10 ng of the sample DNA. The therma
94 °C for 3 min, 35 cycles of 94 °C for 45 s, 50 °C for 1 min, 72 °C for 1.5
min, followed by a 10-min final extension at 72 °C. We used the primer
set ITS1f (Gardes and Bruns, 1993) and ITS2 (White et al., 1990), slightly

l was

(10 mM), 0.50 µl BSA (2%), 1.50 µl of reverse and for-
mers (10 µM) and 0.4 µl Taq polymerase (Invitrogen,
CA). We used the following PCR conditions: an initial
turation step at 95 °C for 5 min; then 35 cycles of 95 °C for 20 s,
54 °C for 30 s and 72 °C for 1.5 min; and ending with one cycle of
72 °C for 10 min. The ITS2 rDNA region was amplified using the

Still several protocols are not optimized to minimize the risk of biases

Sometimes n° of cycles can be as much as 45…



polymerase:DyNAZyme II DNA polymerase, 1:24) and 1 µ
conditions were 94 °C for 5 min, 35 cycles of 94 °C for 1 n
for 30 s, followed by 72 °C for 10 min for primers eub530f/
5 min; 35 cycles of 94 °C for 1 min, 55 °C for 1 m

0.2 µl of 0.056 U fast StartExp*Taq* Polymerase (Roche Applied Sciences, Indianapolis, IN, USA), 2.5 µl dNTP (2 mM each), 0.25 µl of each primer, and 1.0 µl of DNA template. Thermocycling conditions were as follows: denaturing at 95 °C for 5 min followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 53 °C for 30 s, extension at 72 °C for 60 s followed by a final extension at 72 °C for 10 min. As negative control, water was used instead of

ecular grade water, 0.5 µl of 10 µm of the 515F, 0.5 µl of 10 µm 806R
The PCR conditions were the following: 94°C for 5 min, followed by 35
llowed by a final elongation of 10 min at 72°C. The composition of the
. The PCR master mix contained 17.1 µl molecular grade water, 1 µl 5 µM

primer, 0.5 µM of the gITS7 primer, and 1U of 1× Phusion High Fidelity DNA Polymerase (New England Biolabs). Thermocycling conditions were as follows: 98 °C for 30 s, 35 cycles of 98 °C for 10 s, 56 °C for 30 s, 72 °C for 15 s and a final extension at 72 °C for 7 min. Negative controls, consisting of sterile water in place of template DNA, did not yield

ately 250 bp) of the nuclear ribosomal rDNA repeat, using the following
one cycle of 95°C for 5 min; then 37 cycles of 95°C for 20 s, 56°C for 30 s and
ending with one cycle of 72°C for 7 min. ITS is the universal DNA
ngi and has been used in a wide variety of taxonomic and ecological

primer and 10 ng of the sample DNA. The thermal
94 °C for 3 min, 35 cycles of 94 °C for 45 s, 50 °C for 1 min, 72 °C for 1.5
min, followed by a 10-min final extension at 72 °C. We used the primer
set ITS1f (Gardes and Bruns, 1993) and ITS2 (White et al., 1990), slightly

l was

mM), 0.50 µl BSA (2%), 1.50 µl of reverse and for-
mers (10 µM) and 0.4 µl Taq polymerase (Invitrogen,
CA). We used the following PCR conditions: an initial
duration step at 95 °C for 5 min; then 35 cycles of 95 °C for 20 s,
54 °C for 30 s and 72 °C for 1.5 min; and ending with one cycle of
72 °C for 10 min. The ITS2 rDNA region was amplified using the

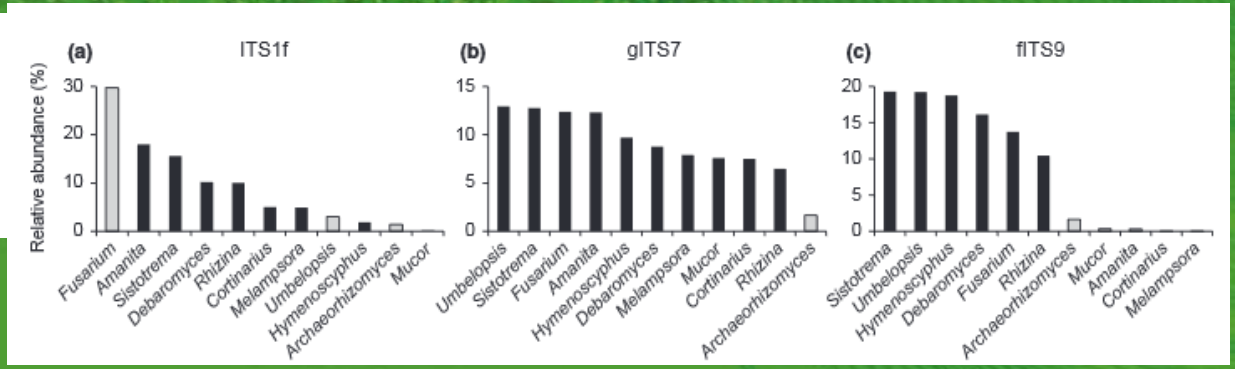
Protocol optimization: Primers, PCR conditions (…)

RESEARCH ARTICLE

New primers to amplify the fungal ITS2 region – evaluation by 454-sequencing of artificial and natural communities

Katarina Ihrmark, Inga T.M. Bödeker, Karelyn Cruz-Martinez, Hanna Friberg, Ariana Kubartova, Jessica Schenck, Ylva Strid, Jan Stenlid, Mikael Brandström-Durling, Karina E. Clemmensen & Björn D. Lindahl

Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden



- Interspecific variation in fungal ITS length
- f(g)ITS7-ITS4 primers showed more or less even amplification of selected group of taxa! But still with some deviations from expected abundances
- Possibility to identify and predict these biases? (And correct for them?)
- Optimization of PCR conditions to minimize biases….
- ….and the use of the less-biased sequencing platform (MiSeq?, IonTorrent?, PacBio?)

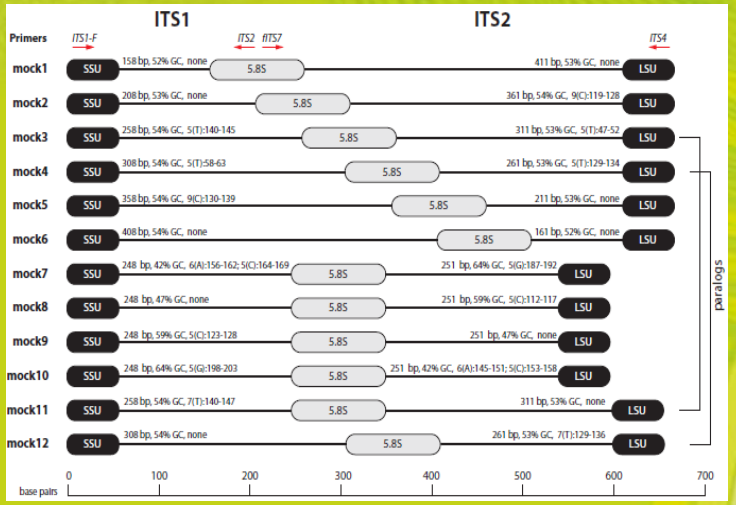
Objectives

- to study the effect of DNA sequencing on the community composition and assess associated sequencing biases
- to quantify and compare biases introduced by three sequencing platforms (Illumina MiSeq, PacBio RS II, PacBio Sequel)
- to assess combined PCR and HTS biases on the final community
- to elaborate a ready-to-use protocol that allows semi-quantitative analysis of fungal communities

Mock communities

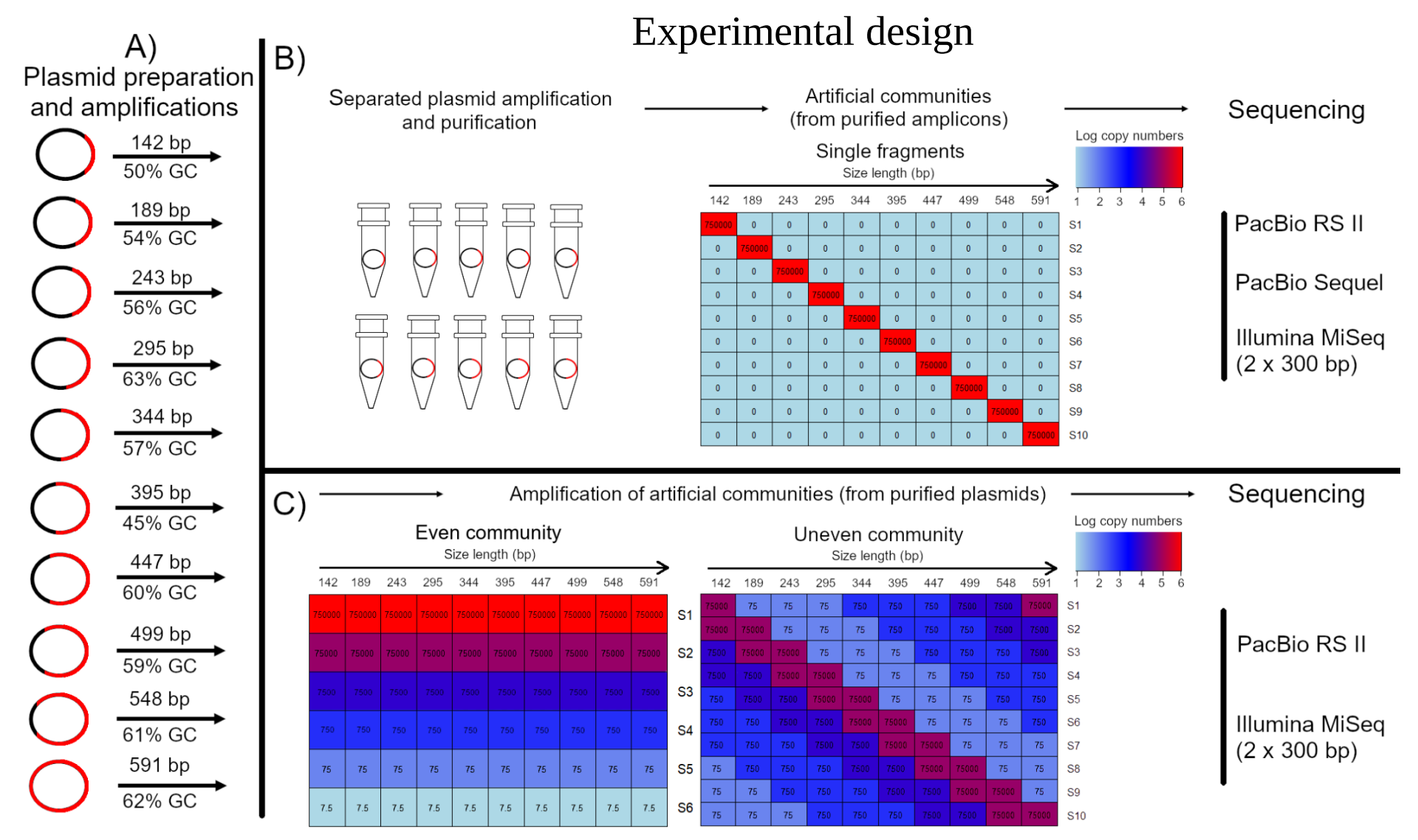
- Excellent tool to assess biases
- Spike-in mock community (ITS MOCK, 10 fragments differing in size length and GC content)

  - ➢ Fragments obtained from *H. annosum* genome
  - ➢ Length: From 180 (min) to 630 bp (max)
  - ➢ GC content: From 45% to 63%
  - ➢ Primers fITS7-ITS4

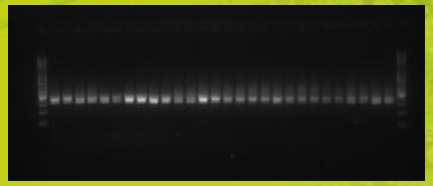


…Fragments extracted from plasmids were combined using distinct assemblies to test the sequencing biases, for PCR and sequencing biases altogether and for PCR biases alone

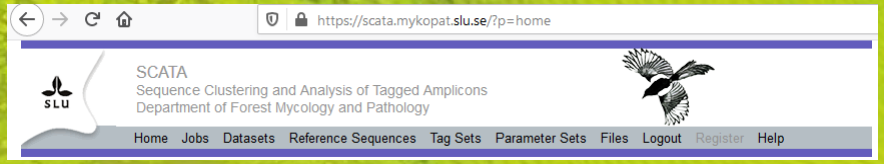
Experimental design



Molecular works and bioinformatics



SCATA pipeline: http://scata.mykopat.slu.se



Molecular works

- fITS7 and ITS4 primers (Ihrmark et al. 2012)
- Both primers tagged using previously validated tags
- Use of distinct PCR cycles
- Purification
- Equimolar mix
- Sequencing adaptor ligated

Quality control

PacBio & Illumina:

- Seq < 100 bp removed
- Screening of primers and tags
- Trimming of sequences with low quality

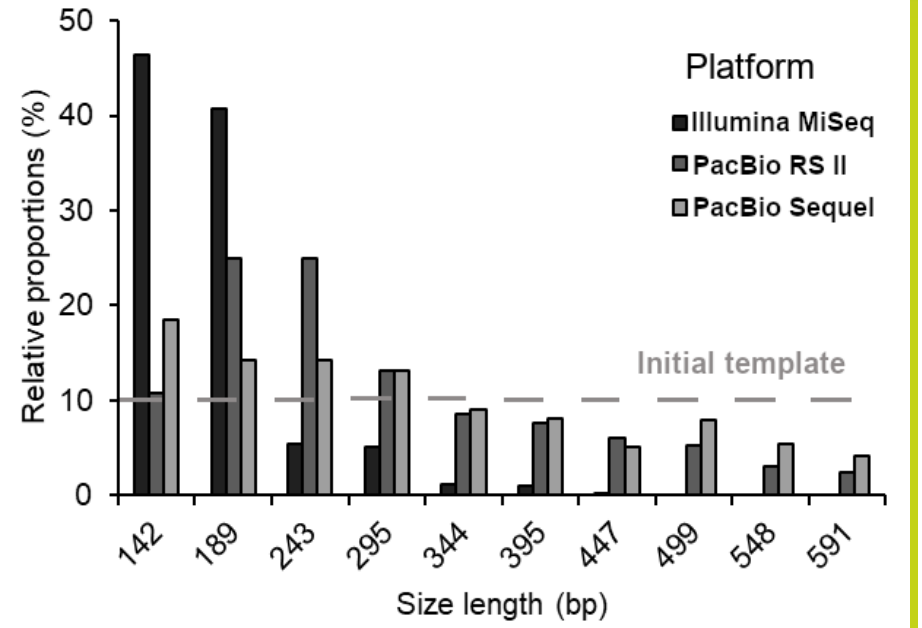
Alignments and sequence clustering

- Usearch (Edgar, 2011)
- Genotypes occurring only once in the global data removed
- Single linkage clustering (98.5% threshold)
- LULU post-clustering algorithm to merge daughter OTUs with parent OTUs (MiSeq)

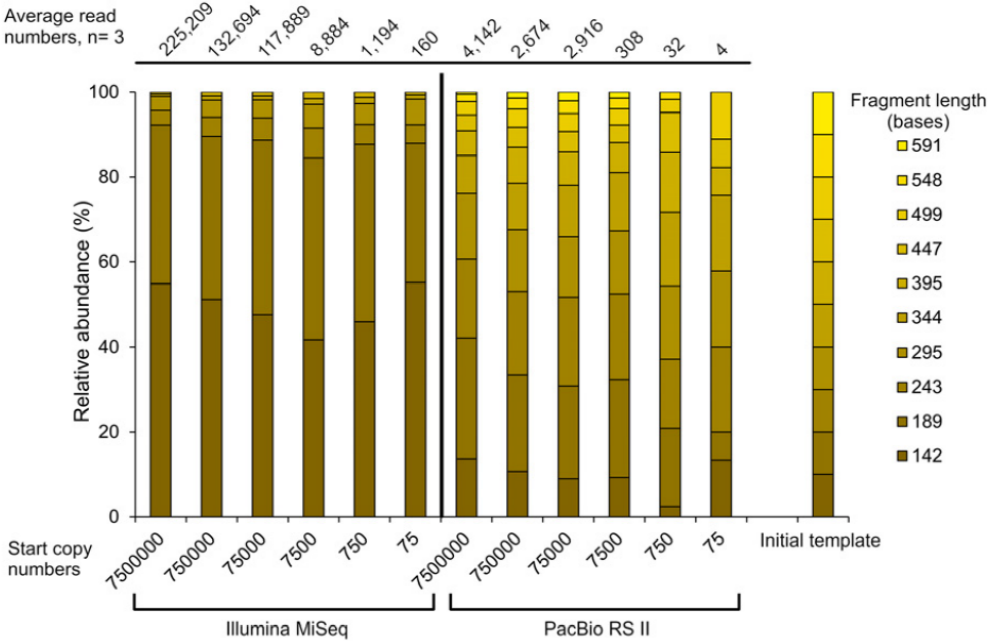
Sequencing biases from evenly mixed fragments

- Sequencing biases largely determined by size length of the fragments, with MiSeq having the highest biases
- PacBio Sequel was the less biased platform

A)

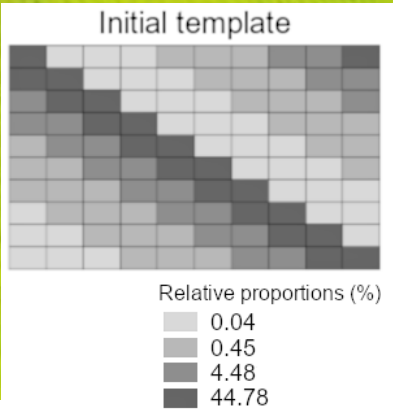


B)



Sequencing and PCR biases from unevenly mixed fragments

- Communities sequenced with PacBio RSII (Sequel not tested here) resembled much more to the original community than Illumina MiSeq did



Deviation (%) from initial template

-40 -20 0 20 40

PacBio RS II

Fragment length (bases)

	142	189	243	295	344	395	447	499	548	591	
S1	38.44	0.12	0.33	0.01	0.77	0.35	-0.08	-1.82	-2.09	-36.02	S1
S2	-17.19	25.64	0.02	0.02	0	-0.05	-0.37	-0.26	-3.7	-4.12	S2
S3	-2.51	7.71	0	-0.04	-0.02	-0.02	-0.35	-0.4	-0.43	-3.94	S3
S4	-1.82	1.75	13.02	-11.84	0.07	-0.04	0.01	-0.4	-0.45	-0.29	S4
S5	-0.11	4.91	6.11	4.1	-14.27	0.02	-0.04	0.07	-0.39	-0.39	S5
S6	0.19	0.94	6.35	2.76	-5.58	-4.14	-0.04	-0.04	-0.04	-0.4	S6
S7	-0.16	1.26	0.67	5.72	3.72	4.15	-15.27	0.01	-0.04	-0.04	S7
S8	0.07	2.6	1.38	1.32	7.43	2.89	-6.94	-8.66	-0.04	-0.04	S8
S9	0.15	0.09	1.3	0.43	0.74	5.71	0.77	3.78	-12.9	-0.04	S9
S10	0.06	0.83	0.44	2.94	1.78	1.88	6.08	1.24	11.52	-26.76	S10

Illumina MiSeq

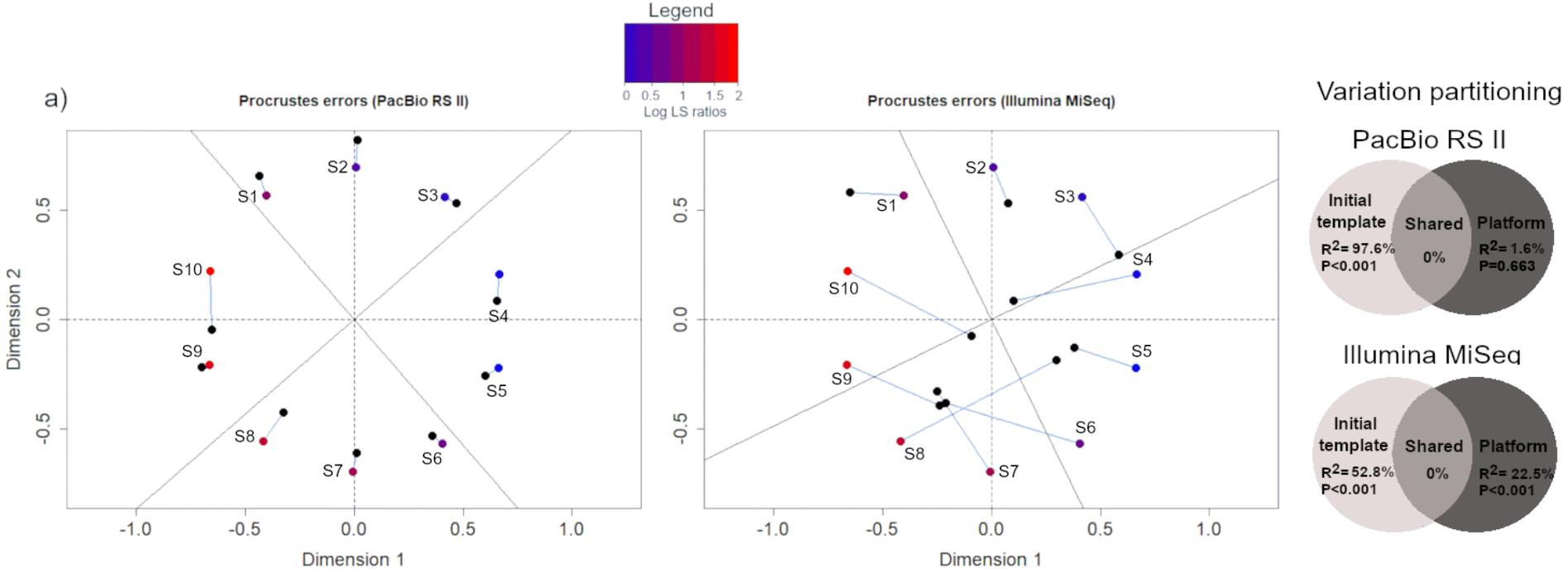
Fragment length (bases)

	142	189	243	295	344	395	447	499	548	591	
S1	55.12	0.06	-0.04	-0.04	-0.45	-0.45	-0.45	-4.48	-4.48	-44.78	S1
S2	9.92	0.52	-0.04	-0.04	-0.04	-0.45	-0.45	-0.45	-4.48	-4.48	S2
S3	4.92	35.92	-34.88	-0.04	-0.04	-0.04	-0.45	-0.45	-0.45	-4.48	S3
S4	23.32	19.42	-16.38	-24.98	-0.04	-0.04	-0.04	-0.45	-0.45	-0.45	S4
S5	4.75	41.82	1.32	-8.78	-38.18	-0.04	-0.04	-0.04	-0.45	-0.45	S5
S6	15.85	12.85	11.32	8.02	-27.68	-19.88	-0.04	-0.04	-0.04	-0.45	S6
S7	21.15	18.15	1.65	14.22	-0.68	-9.68	-44.68	-0.04	-0.04	-0.04	S7
S8	6.16	57.45	6.15	4.55	7.02	7.62	-44.08	-44.78	-0.04	-0.04	S8
S9	17.96	15.46	15.95	13.45	1.75	29.32	-4.38	-44.78	-44.78	-0.04	S9
S10	34.26	25.36	3.76	25.45	5.15	4.55	-4.38	-4.48	-44.78	-44.78	S10

Sequencing and PCR biases from unevenly mixed fragments

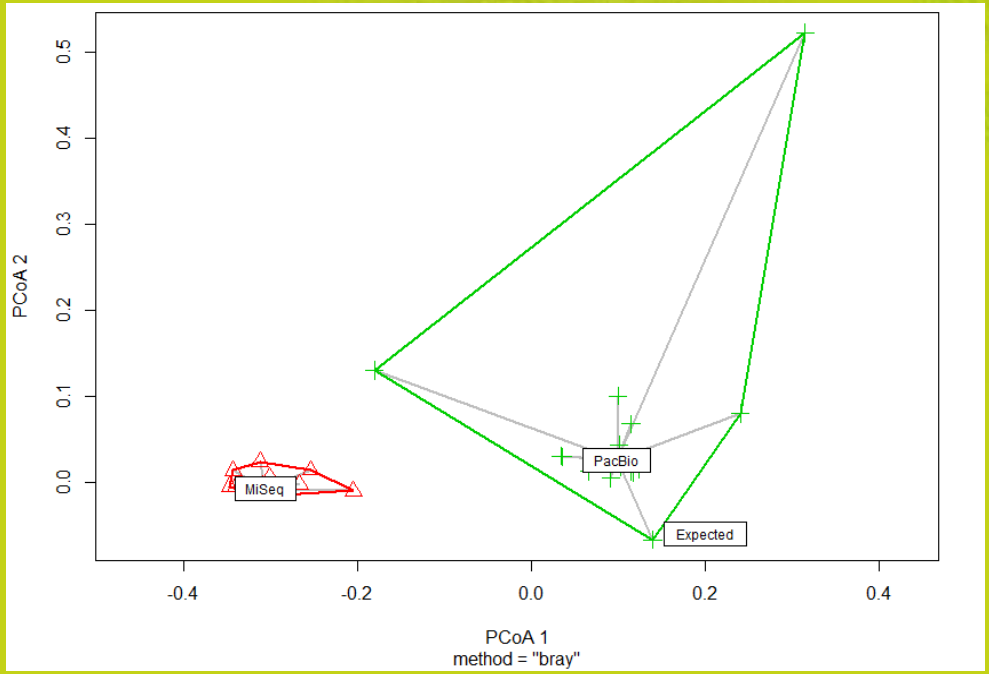
- Communities sequenced with PacBio RSII (Sequel not tested here) resembled more to the original community than Illumina MiSeq did

- MiSeq data also showed high degree of stochasticity or non-explained variation



Sequencing depth

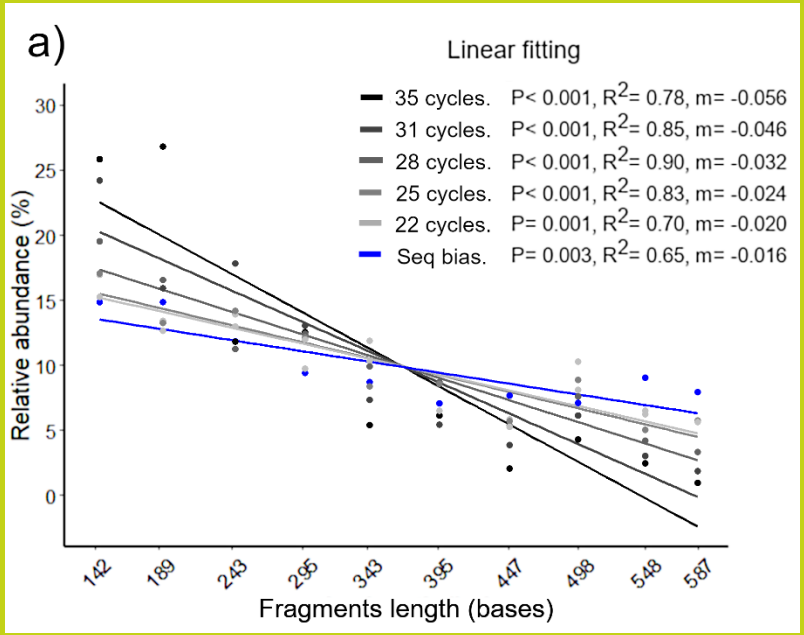
- PacBio Sequel seem to be the best and less biased platform, but be aware of low sequencing depth….
- Highest multivariate variance (beta diversity) in samples sequenced with PacBio, despite these resembled more to the initial composition (Before PCR and sequencing)



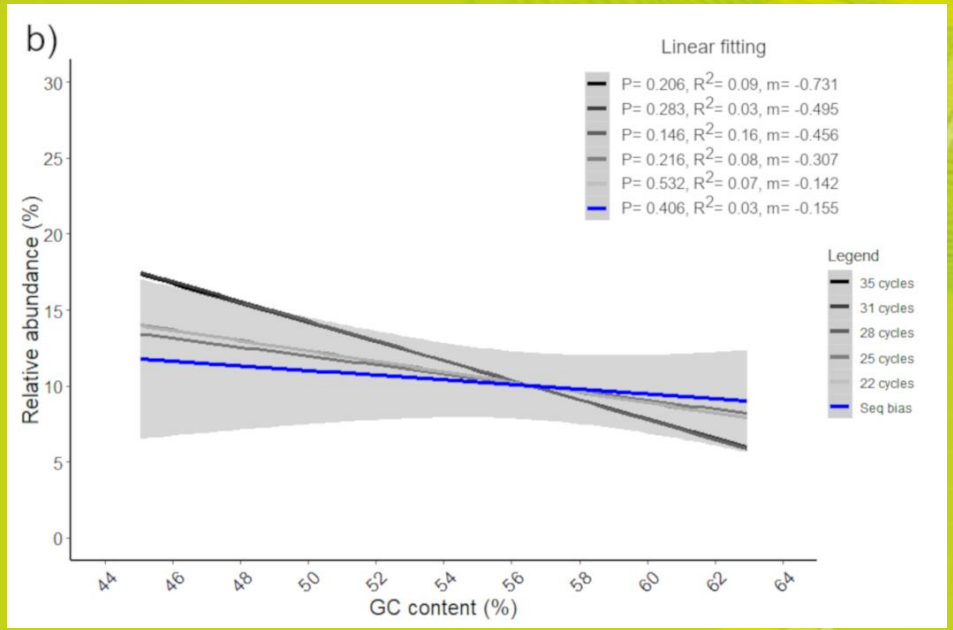
PCR biases alone

- Increasing PCR cycles exacerbate size length biases, but GC content has no influence
- Shortest fragments are over-represented with more PCR cycles, and longest fragments are under-represented

Modelling PCR cycles effect on % of each fragment across distinct lengths



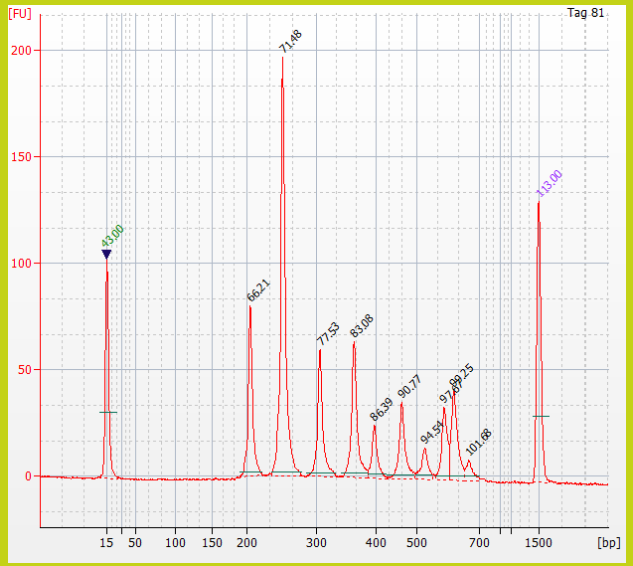
Modelling PCR cycles effect on % of each fragment across distinct GC content



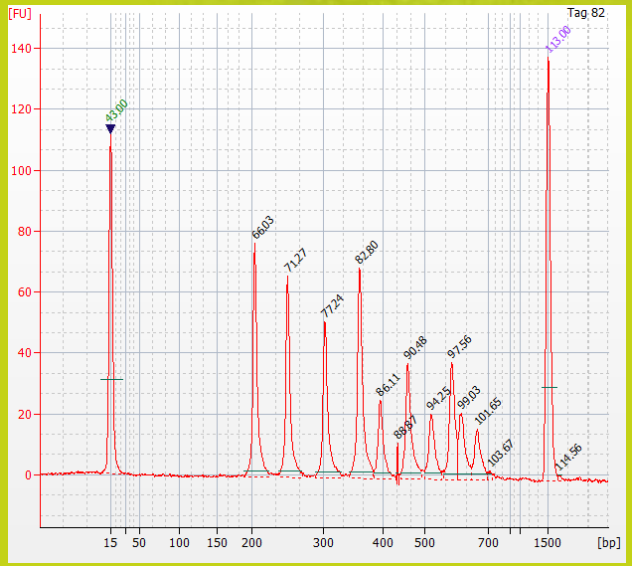
PCR biases alone

- Fragment size biases increased with increasing PCR cycles. Already obvious in BioAnalyzer

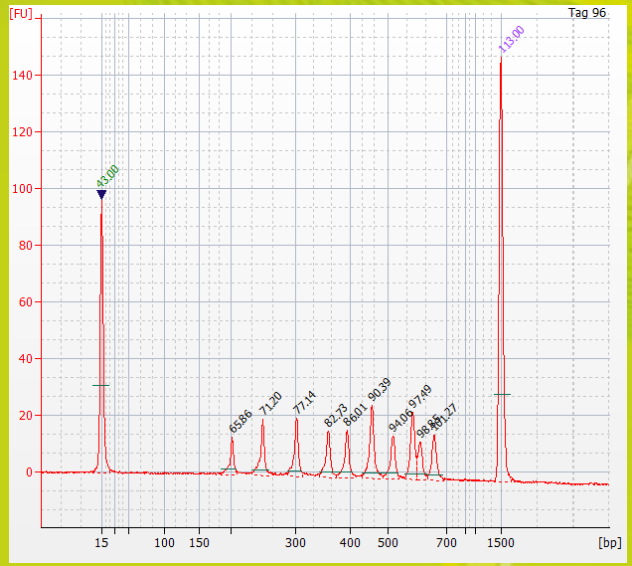
35 cycles



28 cycles



22 cycles

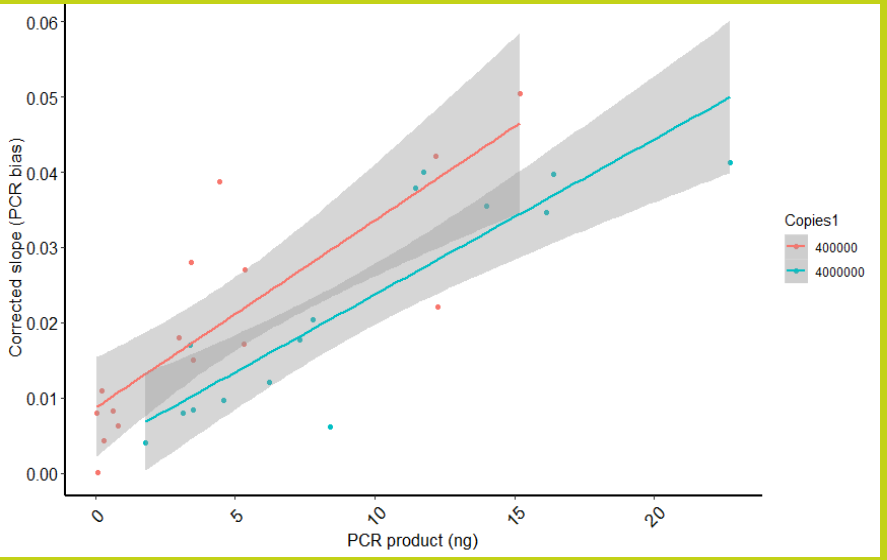


PCR biases alone

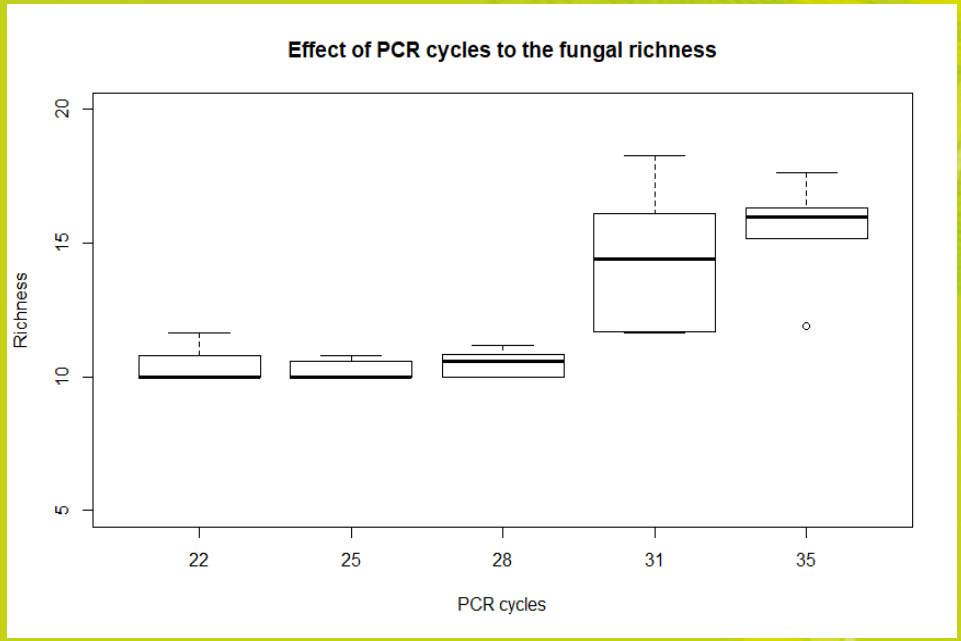
- Final PCR product may be used to predict PCR biases, but overall, PCR cycles predict better biases than final PCR product

- Higher starting quantities allow the user to use less PCR cycles, but there was positive relationship between starting quantities and sequencing errors.

Relationship between PCR product and corrected slope (size length bias)



Effect of PCR cycles to the fungal richness



Conclusions

- (Semi)quantitative analysis of sequencing data is possible; also, possible to account for size length biases
- Length bias is largely the main factor determining total biases during sequencing and PCR
- PacBio Sequel is the sequencing platform that resembled more to the expected communities if biases would not occur
- Biases during PCR can be minimized using low PCR cycles; technically no biases occurring during PCR
- Non optimized PCR conditions exacerbate biases (size length biases, community distortion and errors), but the choice of the sequencing platform is more important to minimize biases
- Low sequencing output may result in stochastic results promoting community dispersion in ecologically similar samples. However, high sequencing depth in Illumina platforms do not alleviate biases
- The message is the need to improve current technologies

New Phytologist

Research

Methods

Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities

Carles Castaño[1], Anna Berlin[1], Mikael Brandström Durling[1], Katharina Ihrmark[1], Björn D. Lindahl[2], Jan Stenlid[1], Karina E. Clemmensen[1]* and Ake Olson[1]*

[1]Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala SE-75007, Sweden; [2]Department of Soil and Environment, Swedish University of Agricultural Sciences, Uppsala SE-75007, Sweden

SCIENCE AND EDUCATION FOR SUSTAINABLE LIFE

SCIENCE AND EDUCATION FOR SUSTAINABLE LIFE